Paper Reviews

Arka Sadhu

July 27, 2017

Contents

1 Unsupervised Co-segmentation for Indefinite Number of Common Foreground Objects 2

1.1 Abstract 2
1.2 Introduction 2
1.3 Problem Formulation 2
1.4 Co-segmentation for Indefinite Number of Common Targets 3
1.4.1 Overall Framework 3
1.4.2 Object Proposals Generation 3
1.4.3 Weighted Graph Construction 3
1.4.4 Common Targets Multi-Search Strategy 4

2 Video Object Co-segmentation by Regulated Maximum Weight Cliques 4

2.1 Abstract 4
2.2 Introduction and Related Work 4
2.3 Regulated Maximum Weight Clique based Video Co-segmentation 5
2.3.1 Framework 5
2.3.2 Object Tracklets Generation 5
2.3.3 Multiple Object Co-segmentation by [RMWC] 5

3 Object-Based Multiple Foreground Video Co-Segmentation via Multi-State Selection Graph 6

1 Unsupervised Co-segmentation for Indefinite Number of Common Foreground Objects

[1]

1.1 Abstract

- Co-segmentation addresses the problem of simultaneously extracting the common targets appeared in Multiple images.

- Keywords : Co-segmentation, multi-object discovery, adaptive feature, loopy belief propagation

1.2 Introduction

The paper extends the previous proposal Selection based Co-segmentation[PSCS] methods with the 3 major contributions :

- Key problem in [PSCS] is mining consistent information shared by the common targets. May require manual selection of features, or feature learning performed beforehand. Here : (simple and effective) self-adaptive feature selection strategy is introduced.
- Many assume each image contains a single common target and fail for multiple common targets images to extract all targets. Here proposal selection based Unsupervised Co-segmentation [PSUCS] is introduced.
- For multiple common targets, multi-class co-segmentation approaches do not do so well because of significant appearance variance and the inconsistent number of common targets, also some combinational common trgets are usually splitted into multiple pieces. Here : an adaptive strategy that can handle Indefinite number of common targets involved cases, where each image may contain different number of common targets.

1.3 Problem Formulation

- Image set $I = I_i, i = 1...M$, images may contain different number of targets, goal is to extract all common targets.
- Given I_i generate proposal set $P_i = p_i^k, k = 1...K_i$, set large value for K, to make sure that the object proposal set covers all potential common targets.
- Cos for indefinite number of common targets is transformed into a labeling problem, given p_i^k, $x_i^k = 1$ for foreground, else 0.
- Union set is viewed as final segmentation result

$$R_i = \bigcup \{p_i^k | x_i^k, k \leq K_i\}$$

- Here, the labeling problem in a completely connected network., where each object proposal [OP] is a node, and connected with weighted edges.
- Multiple OP of each image is conducted separately, but closely related to other images of the collection.
- For each image I_i, choose a proposal in every selecting loop to be the real foreground, and in choosing the new loop, we remove the node of the previous proposal to make sure this new proposal would be considered a target, whether this will be chosen as a target depends totally on the labels of other images.

- **Therefore**, segmentation problem of image of I_i finally becomes finding an optimal labeling set $x_i = \{x_i^k | k = 1...K_i; x_k \in \{0,1\}\}$ by max the energy function (refer to the paper) : function of weights of the edges and some other constraints.
- Weight is non-zero and numerically equal to a similarity score between the proposals (to be introduced later). The constraints mean, for each image only one proposal could be selected per loop, and every proposal in image I_i can be selected only once throughout the selection procedure.
- The formulation is based on the fact that common targets have same characters, and maximizing overall similarity with additional constraint we can make sure newly chosen object proposal is the most similar one to the chosen proposal of the other image. This can be solved via greedy optimization.

1.4 Co-segmentation for Indefinite Number of Common Targets

Key is to adaptively determining the number of targets, which require fully extracting teh potential targets and then mining the conistent relationships shared by teh common targets.

1.4.1 Overall Framework

1. Category independent OP generated.
2. Connected graph with all proposals as nodes and edge weights as proposal similarities.
3. For reliable similarity, adaptive feature weight selection algorithm.
4. Multiple common targets [MCT] searching, where [MCT] are extracted for each individual image.
5. Terminal condition designed as the common target judging criterion.
6. After termination, simply collect selected proposals.

1.4.2 Object Proposals Generation

1. Very important, directly impacts the performance of Co-segmentation.
2. Measurement of the proposal pool contains mainly two aspects :
   - Diversity : cover as many objects as possible.
   - Representativeness : as few candidates as possible for each object.
3. After a large number of proposals are achieved, a scoring mechanism that combines appearance features and overlap penalty is raised for proposal ranking. There is problem of the proposal containing a local part, but the proposed method could make up for such loss by conducting multiple targets searching.

1.4.3 Weighted Graph Construction

- Usual way : measuring similarity between every two proposals.
- Choosing fixed features for similarity is not a good option. Adopting a flexible and reliable proposal similarity measurement. Here : Unsupervised self-adaptive similarity measurement is introduced for calculating edge weights. Highly efficient and easy to implement. Example in two images colors might be same, in two other images color might be drastically different.
- Use iterative weights setting mechanism for the features. Initial proposal labels using loopy belief algo previously, and then iterating to maximize a function.

- The intuitive intention : encourage selected common targets to be globally consistent while keping a low variance to make the similarity metric more reasonable and representative.

1.4.4 Common Targets Multi-Search Strategy

Adaptive common target searching strategy that can deal with any numbers of targets.

- For more common targets , remove previously discovered ones from the candidate pool.
- Initialize labels x^* from prev algo. Basically selecting most likely common targets, by removing the prev most likely common target.
- Get an adaptive threshold.

2 Video Object Co-segmentation by Regulated Maximum Weight Cliques

[2]

2.1 Abstract

- Novel approach for object co-segmentation in arbitrary videos by sampling, tracking, and matching [OP] via a Regulated Maximum Weight Clique [RMWC] extraction scheme.
- Achieves good results by pruning away noisy segments in video through selection of [OP] tracklets athat are spatially salient and temporally consistent, and by iteratively extracting weighted groupings of objects with similar shape and appearance (with-in and across videos).
- Approach is general and handles : multiple objects, temporary occlusions, objects going in and out of view, also doesn't make any prior assumption on the commonality of the objects in the video collection.
- Keywords : Video Segmentation, Co-segmentation

2.2 Introduction and Related Work

Goal is to discover and segment objects from a video collection in an unsupervised manner.

- Video Co-segmentation is natural extension of Image Co-segmentation.
- In general for video cos, appearance info to group pixels in a spatio-temporal grpah and/or employ motion segmentation techniques to separate objects by using motion cues.
- Previous work use strong assumptions of single class of object common to all videos.
- The work has the following advantages :
    - Employs object tracklets as opposed to pixel-level or region-level to perform clustering. The eprceptual grouping of pixels before matching reduces segement fragmentation and leads to a simpler matching problem.
    - No approximate solution. [RMWC] has an optimal solution. Using only object tracklets keeps the computation cost low.
    - Can handle occlusions, or objects going in and out of the video because the object tracklets are temporally local and there is no requirements for the object to continuously remain in the field of view of the video. Also no limit on teh number of object classes in the each video and number of common object clases in teh video collection. Therefore more general.

- Different from [MWC], in that it is regulated by intra-clique consistency term, as a result produces more global consistency.

2.3 Regulated Maximum Weight Clique based Video Co-segmentation

2.3.1 Framework

Two stages :

1. Object Tracklet Generation : generate [OP] for each frame and use each of them as a starting point and track the object proposals backward and forward trhoughout the whole video seq, and generate reliable tracklets from the track set and perform non-maxima suppression to remove noisy or overlapping proposals.
2. Multiple Objects Co-segmentation by Regulated Maximum Weight Cliques : Tracklets as node, and the nodes are weighted by tracklet similarity, and edges with weight below a threshold are removed. [RMWC] to find objects ranked by score which is a combination of intra-group consistency and Video Object scores.

2.3.2 Object Tracklets Generation

- Generate a number of [OP]. Each proposal has a Video Object Score : combination of motion and appearance.
$$S^{object}(x) = A(x) + M(x)$$
- $A(x)$: appearance score described directly by algo. High for regions with closed boundary in space, different appearance from its surroundings and is salient.
- $M(x)$: motion score defined as the average frob norm of optical flow gradient around the boundary of object proposal.
- Efficient Object Proposal Tracking :
  - Track every object proposal from each frame backward and forwar to form a number of tracks for the object.
  - Combined color (color histograms to model appearance)+ location(overlap ratio) + shape similarity (contour of region in normalized polar coordinates and sampling it from 0 - 360 deg to form a vector) and then dot product for the first and last.
  - Greedy tracking : most similar object proposal is selected to be tracked down, computationally requires fnding index of max value in a specific row of the similarity matrix and hence economical.
- Non-maximum Suppression [NMS] for Object Proposal Tracks :
  - Need to prune duplicate (near-duplicate) tracks.
  - Video Object score for one track is obtained, and see $R_{overlap} > 0.5$ and remove them.
  - After [NMS] small percentage of total tracks are retained, and to ensure validity of the track associations, remove associations that are 1.5 std from the mean track similarity.

2.3.3 Multiple Object Co-segmentation by [RMWC]

After object tracklets are obtained, need salient object groupings in the video collection. Grouping problem is formulated as Regulated Maximum Weight Clique.

- Clique Problems :
  - Given G = (V,E,W), a clique is complete subgraph of G, i.e. one whose vertices are pairwise adjacent.

- Maximal Clique is a complete subgraph not contained in any other complete subgraph.
- Finding all maximal Clique is NP-hard. Maximum Clique problem is to find the Maximum complete subgraph and Maximum Weight Clique problem deals with finding the Clique with max weight.

- Problem Constraints :
  - Object Proposal Tracklets [OPT] : similar appearance both in video and across video, for in-video L channel used, for across a,b also used.
  - Shape of same object would not change in the same video, and hence used for building tracklets of same objects in a video.
  - Dominant object => high Video Object Score [VOS]
  - Tracklets generated by an object should have low variation.
- Graph Structure :
  - Object tracklets [OT] are nodes, inter and intra video edges created as described above.
  - Weak edges removed by a threshold.
- RMWC :
  - Get weight of node.
  - According to formulation : Clique that has the highest score represents the object with largest combined score of inter-object consistency and objectness. Use NP hard formulation, but doesn't hinder its usage, as number of tracklets are limited, and takes less than a second on standard laptop.

3 Object-Based Multiple Foreground Video Co-Segmentation via Multi-State Selection Graph

[3]

References

[1] K. Li, J. Zhang, and W. Tao, "Unsupervised co-segmentation for indefinite number of common foreground objects," *IEEE Transactions on Image Processing*, vol. 25, pp. 1898–1909, April 2016.

[2] D. Zhang, O. Javed, and M. Shah, *Video Object Co-segmentation by Regulated Maximum Weight Cliques*, pp. 551–566. Cham: Springer International Publishing, 2014.

[3] H. Fu, D. Xu, B. Zhang, S. Lin, and R. K. Ward, "Object-based multiple foreground video co-segmentation via multi-state selection graph," *IEEE Transactions on Image Processing*, vol. 24, pp. 3415–3424, Nov 2015.